WPS3362

Education, Information, and Smoking Decisions

Evidence from Smoking Histories, 1940-2000

Damien de Walque

Development Research Group, World Bank

ddewalque@worldbank.org

World Bank Policy Research Working Paper 3362, July 2004

The author thanks Gary Becker, Mark Duggan, Michael Greenstone, Steven Levitt and Tomas Philipson for useful comments, suggestions and encouragements. I am grateful to Daniel Parent for suggesting the Vietnam draft as an instrumental variable for college education. I also thank Raphael De Coninck, Michael Grossman, Ted Joyce, Fabian Lange, Ellen Meara, Chris Rohlfs and participants of the Applications Workshop (May 5, 2003) at the University of Chicago, as well as participants at a NBER Health economics seminar for helpful comments and discussions.

Abstract

This paper tests the hypothesis that education improves health and increases people's life expectancy. Smoking histories—reconstructed from retrospective data in the National Health Interview Surveys in the United States—show that after 1950, when information about the dangers associated with tobacco consumption started to diffuse, the prevalence of smoking declined earlier and most dramatically for college graduates. More educated individuals are also more likely to quit smoking: incidence analysis of smoking cessation shows a strong education effect. The instrumental variable approach, which relies on the fact that during the Vietnam War college attendance provided a strategy to avoid the draft, indicates that education does affect decisions about whether to smoke or stop smoking.

1. Introduction

The strong correlation between education category and health outcomes, even after controlling for income, has been recognized as a robust empirical observation in the social sciences and economic literature (Deaton and Paxson 2003; Fuchs 1982; Grossman 2004; Lleras-Muney 2004). The decision to smoke or not to smoke is a conscious choice that directly affects the health status and ultimately the mortality of individuals. It therefore provides an interesting opportunity to investigate how education, by influencing behaviors, affects health outcomes.

Smoking habits were not initially perceived as dangerous. The information revealing the health damaging consequences of cigarette smoking emerged gradually between 1950 and 1970. Variations in smoking prevalence across education groups during this period might inform us about the way individuals reacted to that information and how education is instrumental in accessing and processing this information.

It turns out smoking is the leading cause of premature adult mortality. Each year in the United States, tobacco use causes more than 440,000 deaths and results in more than 5.6 million years of potential life lost.[1] The impact on mortality of smoking, and the potential role of education in prevention suggest that traditional estimates of the returns to education—focusing exclusively on labor market outcomes—might be too low. This paper examines whether the effect of education on improving health outcomes can be considered as causal.

This issue is central in the economics literature on the interaction between health and education. Theoretical explanations for this correlation can be classified into three broad categories. One explanation stresses that education is an investment. Education will deliver a higher income, a higher consumption level in the future, and raises the value of staying alive (Becker 1993). More educated individuals are healthier because their investment in the future gives them the right incentives to protect their health. Another explanation, based on education entering as a factor in the health production function (Grossman 1972), emphasizes that education improves the access to health-related information and the processing of that information to make health-related decisions. A third view (Farrell and Fuchs 1986; Fuchs 1982) claims that the observed correlation between health and education is mainly due to unobservables, like the discount factor or the ability that causes the same individuals both to study longer and to take greater care of their health.

This paper analyzes smoking prevalence across education groups in the United States from 1940 to 2000. Despite the lack of surveys linking education and smoking

[1] http://www.cdc.gov/tobacco/issue.htm (accessed on May 11, 2004).

before 1966, this information is obtained by using retrospective smoking histories constructed from the smoking supplements of National Health Interview Surveys conducted between 1978 and 2000. The conclusion from this analysis is that the smoking prevalence among more educated individuals, college graduates in particular, declined earlier and most dramatically than any other education category. The decline for college graduates started in 1954, four years after the medical consensus on the health consequences of smoking and 10 years before the publication of the first Surgeon General's Report on this issue. This suggests that they had easier access to the information and/or were more able to process that information.

The decision to quit smoking—usually taken after education is completed—provides a clearer margin for analysis than smoking initiation, which generally occurs before schooling levels are definitive. The paper further develops a detailed study of the incidence of smoking cessation using a proportional hazard model, which shows the strong effect of education on the probability for smokers to quit smoking at any point in time. The analysis for all birth cohorts of smoking age during the 1950–70 period, when the information on the risks associated with tobacco was not uniform, indicates that it was during this period that the effect of education was the strongest, independent of age. The instrumental variable approach used—college attendance as a draft avoidance strategy during the Vietnam War—suggests that a college education affects decisions about whether to smoke or stop smoking, even after controlling for income.

The paper is structured as follows. Section 2 presents the results from the analysis of the smoking prevalence in the United States from 1940 to 2000. Section 3 focuses on an analysis of the incidence of smoking cessation among smokers. Section 4 contains the results of the instrumental variable strategy that exploits, as an instrument for education, college attendance as a draft avoidance strategy during the Vietnam War. Section 5 concludes.

2. Evidence from Smoking Histories in the United States, 1940–2000

It is well known by social scientists that more educated people are less likely to smoke (Wald 1988). The literature about the negative correlation between education and tobacco use has become a key element in the debate about whether education is causal in improving health.[2]

The study of smoking habits and education prior to the arrival of the information about the dangers of smoking for the health have been less studied due to lack of data (Ferrence 1989; Harris 1983; United States Government 1980). Before 1950, even though

[2] See Berger and Leigh (1989); Chaloupka and Warner (2000); Currie and Moretti (2003); Farrell and Fuchs (1986); Fuchs (1982); Grossman (1975); Kenkel (1991); Meara (2001); Sander (1995); Viscusi (1990, 1992).

the medical literature mentioned tobacco smoking as a possible factor in causing mouth and lip cancer, there was no consensus in the medical profession. This consensus was achieved in 1950, with the publication of five studies, four in the United States and one in the United Kingdom showing the link between smoking and lung cancer (Peto and others 2000). These medical findings were echoed in the popular press in 1953 in the ‘‘Reader’s Digest’’ and in ‘‘Consumer Report’’ (United States Government 1989; Cutler and Kadiyala 2003; Viscusi 1992). In 1964, the publication of the first Surgeon General’s Report (United States Government 1964) gives the endorsement of the government and disseminates the information to a larger public. The release of this report has been ranked one the most important news event of 1964 (United States Government 2004). It marks the beginning of awareness campaigns and prevention efforts. In 1966, a first health warning was made mandatory on cigarettes packages. The wording was rather timid: ‘‘*Caution: Cigarette Smoking May be Hazardous to Your Health’’*. Long debates in Washington (National Conference on Smoking and Health 1970) brought in a stronger language in 1970*:* “*Warning: The Surgeon General Has Determined that Cigarette Smoking is Dangerous to Your Health.*” Given the overwhelming majority of literate smokers in 1970, the health damaging consequences of smoking can be considered as common knowledge by then. The period 1950-1970 can thus be interpreted as a period of gradual diffusion of that information and it is reasonable to expect the more educated individuals to have had access to that information earlier in that period. How did this differential access to information impact smoking behaviors? This part of the paper will address this question.

2.1 Data

The problem in studying smoking behaviors before and after the health damaging consequences of smoking became widely public is the lack of data linking smoking prevalence and education category before 1966. The first smoking supplement included in the National Health Interview Survey dates from 1965[3], and in 1966 education categories were included (United States Government 2004).[4]

---

[3] Gallup surveys about tobacco use were available as soon as 1949, but do not classify the respondents by education category (Cutler and Kadiyala 2003; Viscusi 1992).

[4] The only statistics taken from current surveys where some information about the relationship between smoking and education levels is available are from the United Kingdom (Wald 1988). In 1948, men from almost all ‘‘social classes’’ were smoking at a very similar and very high level (between 64 and 72 percent). Only the lower ‘‘class’’, i.e., manual workers earning less than 7£ per week were smoking significantly less (45 percent). By 1985, the situation had reversed: the education/smoking gradient was clearly negative (smoking prevalence among skilled and partly skilled manual occupations was 40 percent; 34 percent among skilled manual occupations; 28 percent for clerical and lower professional, and 20 percent for professional clerical occupations). To get information about the level of smoking prevalence in the U.S. population before 1966, most studies rely on sales data to compute the per capita number of cigarettes consumed per year (United States Government 1994). This approach does not allow disaggregating consumption of cigarettes by education categories.

This paper uses 16 smoking supplements from different years of the National Health Interview Survey between 1978 and 2000 to construct smoking histories going back before 1950. A similar method, based on cohort reconstruction, has been used in the sociological and medical literature with U.S. and Canadian data (Ferrence 1989; Harris 1983; United States Government 1980).

Since 1978 the smoking questionnaire in the National Health Interview Survey includes questions about the age at which individuals started smoking cigarettes, if they ever started, and about how long ago they stopped, if they did so. Using the year at which individuals started smoking and the year at which they stopped, I constructed 373,738 smoking histories of adults aged 25 and above at the time of the interview.[5]

To make sure that the education variable is not time sensitive, and given that most people will have reached their definitive education level by age 25, I have sampled only individuals aged 25 and above.

2.2 Smoking Prevalence by Education Category, 1940–2000

Figure 1 shows how the general smoking prevalence evolved in the United States from 1940 to 2000. The years 1950 (consensus in medical journals), 1964 (first Surgeon General Report) and 1970 (clear health warning on the packages) are benchmarks in the gradual diffusion of the information about the health consequences of tobacco. Smoking prevalence in the population aged 25 and above was 37.7 percent in 1940 and reached a peak of 45.9 percent in 1957, before the publication of the Surgeon General's report. After that, it steadily decreased and reached 24.6 percent in 2000.

Figure 2 breaks down the data used in figure 1 in four educational groups: individuals with less than a high school degree, high school graduates, individuals with some undergraduate education and finally college graduates together with individuals having studied at the graduate level. It is worth mentioning that each point on figure 2 is estimated from at least 1,000 observations. In 1940 individuals with less than a high school degree are the least likely to smoke (35.8 percent). The smoking prevalence for the three other education categories is all close to 40 percent (39.4, 40.8, and 40.4 respectively). By 2000 there is a clear negative gradient between educational achievement and smoking

[5] To estimate the smoking prevalence in any year, I created a dummy variable for every individual who interviewed after or during that year and was aged 25 and above in that year. The dummy takes the value zero if the individual never smoked, started after that year or stopped before it. It takes the value 1 if the individual smoked during that year. In the 1983, 1985, 1990, 1991 and 1994 surveys, no questions about the age at which individuals started smoking were included. Given that, in the other 11 surveys, 93.3 percent of the respondents who ever started smoking had started by age 25, and since I only sampled individuals aged 25 and above, I considered every person to have ever smoked in these five surveys has having started by age 25. Individuals might stop smoking for a certain period and then resume. A few surveys (1978, 1979, 1980, 1990, 1991, 1992) included questions about temporary smoking cessation, but since in 86.2 percent of the cases the period of temporary smoking cessation was not longer than one year, I have for simplicity ignored it and decided to use only the starting year and the year of definitive cessation.

prevalence. The prevalence of smoking among individuals with less than a high school degree is 29.6 percent, 28.4 percent for high school graduates, 25.6 percent for people with some college education, but only 14.2 percent of people with at least a college degree smoke. The most striking feature of figure 2 is that smoking prevalence among the college graduates declined earlier and more dramatically. This declining trend starts in 1954, 10 years before the first Surgeon's General Report and only one year after the first articles in the general press. Differences in smoking prevalence among the three lower education categories are less marked, but a careful examination of figure 2 reveals, however, that the ranking between these three categories has been inverted between 1940 and 2000. Figure 2 clearly supports the idea that more educated people reacted more quickly and more strongly to the information about the health damaging consequences of smoking.[6]

Keep in mind that in this analysis the general level of education in the population has been steadily increasing between 1940 and 2000. Being a college graduate in 1940 was more exceptional than it was in 2000. Figure 3 addresses this question by plotting the smoking prevalence among individuals above or under the average education level in each year for individuals aged 25 in that particular year.[7] Initially, individuals above the average education level were more likely to smoke. The positive difference between the prevalence in both groups has been gradually decreasing and now the situation has been inverted, since individuals under the average schooling level in the population are much more likely to smoke.

2.3 Sample Composition Issues

The retrospective smoking histories allow a complete overview of smoking prevalence patterns by education levels from 1940.[8] However, this method presents two problems in the composition of the sample. First, since the information about smoking prevalence between 1940 and 1977 is gathered from surveys taken between 1978 and 2000, only individuals who survived up to the year of the survey are interviewed. Since smokers usually die sooner than nonsmokers, this creates a ''survivor'' bias: nonsmokers are disproportionately represented among persons interviewed at older ages. This problem is avoided by selecting respondents less than 60 years of age at the time of the survey. Even though there is already some excess mortality due to smoking before age 60, most of

---

[6] De Walque (2003) confirms that this general picture is valid when males and females are analyzed separately.

[7] In 1940, the average education level was between 10 and 11 years of education, in 1954 it went above 11 and in 1976 it went over 12. This explains the discontinuity in 1976, since in that year high school graduates, who were previously above the average, fell under the average.

[8] I have compared current smoking prevalence rates by education from 1974 to 1999 as calculated directly from the yearly National Health Interview Survey with the estimates obtained in this paper using retrospective smoking histories constructed from 16 surveys. The estimates using these two methods are very close. Results are available on request from the author.

the premature deaths from smoking-related diseases occur between age 60 and 75 (Peto and others 2000). Figure 4 plots the smoking prevalence from 1945 to 2000 by education categories for respondents aged 60 or less at the time of the interview. The historical patterns of smoking prevalence by education levels are similar to the ones described previously.[9]

The fact that data for the period 1940-1977 are taken from interviews after 1977 implies that the age distribution for the years before 1978 is not complete. The data for the year 1940, for example, is made from individuals 25 years of age and above, but young enough in 1940 to have survived up to at least 1978, i.e., 38 years later. As one gets closer to 1978, the age distribution in the sample becomes closer to the age distribution in the general population. To address this problem and maintain a constant age distribution, figure 5 plots the smoking prevalence by education for individuals between the age of 25 and 30 in each year. If we consider that figure 5 is somewhat less precisely estimated given the reduced sample size implied by the selection of the 25-30 age group, figures 4 and 5 are very similar.[10]

Even though the above mentioned sample composition issues, which are inherent to the retrospective nature of the data, need to be taken into consideration and have some influence on the estimation of the smoking prevalence between 1940 and 1950, they do not affect the main conclusion already drawn from figure 2: more educated individuals reacted earlier and more dramatically to the arrival of information about the health consequences of smoking. By controlling for age at the time of the interview and year of birth, multivariate regression analysis of the type performed in sections 3 and 4 also allows addressing some of those sample composition issues.

Another problem, however, with the way the data set has been constructed, is that it is only possible to use time-invariant variables like gender, race, and educational achievement (which I assume does not vary after age 25). Variables such as income or the number of cigarettes smoked per day, present in most of the surveys, are measured at the

---

[9] However, a comparison between figures 2 and 4 confirms the existence of a survivor bias. This is best seen in the 1945-1950 period: in figure 4 the smoking prevalence is higher (above 50 percent) than in figure 2 and the trend up to the early 1960s is rather flat, whereas figure 2 suggests an increasing prevalence up to that point. These two comparisons suggest that in figure 2, between 1940 and 1950, nonsmokers are over-represented because of the survivor bias. By construction, however, the importance of this survivor bias decreases over time.

[10] The two sample composition issues discussed here explain probably also why the less-educated category appeared less likely to smoke in 1940 in figure 2 but not in figures 4 and 5: earlier in the century, before 1940, tobacco might have been too expensive for the less-educated segment of the population, so that they were less likely to start smoking. Older people with low education were not likely to smoke in the 1940s so this dragged down the smoking prevalence for individuals with no high school degree. When the sample is restricted as in figures 4 and 5 to people that were young between 1940 and 1950, the smoking prevalence for the individuals with less than a high school degree between 1940 and 1950 becomes very close to the prevalence of the other education groups.

date of the survey but, since they can vary over time, they are not included in the specifications for which I rely on retrospective data.[11] Section 4, however, by investigating whether or not the individuals are currently smoking as a dependent variable, will allow for controlling for income.

The price of cigarettes is also an important element of the smoking decision, even if tobacco is an addictive good (Becker, Grossman, and Murphy 1994). The National Health Interview Surveys do not include information about prices but the evolution of the real, after-tax price of a pack of 20 cigarettes from 1955 to 2000 is reconstructed in De Walque (2003). Despite a slight increase in the early 1970s, the real price of a pack of cigarettes has been fairly constant up to the early 1980s. Substantial increases in taxation after 1980 and the Tobacco Settlement in the late 1990s explain the rise in the real price to consumers. Figure 2 indicates that most of the differential in smoking prevalence by educational level is already present before the early 1980s and has occurred when there was no substantial variation in the real price of cigarettes.

3. Smoking Initiation and Smoking Cessation

Smoking prevalence in the population is a stock that is defined by two flows: smoking initiation and smoking cessation. The upper panel of table 1 presents the results of a logit estimation where the dependent variable is an indicator for whether the individual ever started smoking. The estimation has been made for all individuals born between 1910 and 1969 and then separately for each 10-year birth cohort in that interval. In order to facilitate the interpretation of the results, I have indicated the age of each birth cohort in 1950 and in 1970, the starting and ending points of the period of gradual introduction of the information about the dangers of smoking. The average of the marginal coefficients, taken observation by observation, of the logit estimation are reported. In order to control for the sample composition issues discussed earlier, survey and year of birth dummies have been included. The other regressors are a set of dummies for the education categories (less than high school is excluded) and dummies for gender and race categories (female and white are excluded).

The results on the whole sample indicate that the likelihood of smoking decreases with schooling. The most striking results of the estimation, however, are found in the evolution of the coefficients on the education categories dummies when progressing across the birth cohorts. For the first two birth cohorts, born between 1910 and 1929, higher educational achievement does not necessarily lead to a lower probability of smoking. These two birth cohorts were aged between 21 and 40 in 1950 and between 41 and 60 in

[11] Using the single 1990 National Health Interview Survey, I verified that considering the quantity of tobacco smoked does not change the conclusion (De Walque 2003). Controlling for income, the higher the level education among smokers, the smaller the amount of cigarettes smoked daily. This result reinforces my previous conclusions: more educated people are less likely to smoke and, if they smoke, they are smoking fewer cigarettes per day.

1970. The information about the dangers of smoking was made available to them after they had taken the decision to start smoking (a decision usually made in the late teens or early twenties). However, individuals with at least a college degree were always found less likely to have started smoking. But the striking pattern is that the magnitude of the negative coefficient on the college and above dummy increases steadily with each birth cohort. For the two other categories, the coefficients, which are positive for early cohorts, also become more and more negative for younger cohorts. With the gradual arrival of the information about the dangers of smoking, the role of education in explaining differences in take-up rates of smoking is larger. It is also noteworthy to realize that white males are more likely to have started and African Americans less likely, but that the effect of those variables is declining for younger birth cohorts.

The lower panel of table 1 contains the results of a similar logit analysis for the other main determinant of smoking prevalence, the decision to quit smoking. The sample size is restricted to the individuals who started smoking. The results point in the same direction as in the upper panel: the probability to have stopped smoking increases with the education level. And again, the striking result is that the coefficients on the education categories are increasing with the year of birth of the cohorts.[12] Interestingly, the results reveal that if females and African Americans were less likely to start smoking, they are less likely to quit if they have ever started. This difference in quitting behavior by gender disappears, however, for the younger cohorts.[13] The conclusion from table 1 is that the differences in the decline in smoking prevalence across education groups are due both to differences in behaviors at the entry and exit points: more educated individuals are less likely to start and more likely to stop.

The distinction between the two margins of the smoking decision—starting vs. quitting smoking—is relevant when considering the relationship between education and smoking. Indeed, the median age of smoking initiation is 17, before most differences in educational achievement take place, while the median age of smoking cessation is 37, when the schooling levels of most individuals can be seen as definitive. Farrell and Fuchs (1986) use the age pattern of smoking initiation to show that "the strong negative relation between schooling and smoking observed at age 24 is accounted for by differences in smoking behavior at age 17, when all subjects were all in the same grade." They use this finding to reject the causality from schooling to smoking, in favor of a "third variable"

[12] The only exception is for the last birth cohort, but this is explained by the fact that those individuals were interviewed at much younger ages and therefore have had less time to decide whether or not to smoke. This is confirmed by the fact that the observed probability to have stopped smoking is declining as the cohorts get younger. Considering this effect of age on the probability of smoking cessation, the fact that the coefficients on the education dummies increase up to the 1950-59 birth cohort is even more striking.

[13] De Walque (2003) also reports the results of a linear regression where the dependent variable is the age at which individuals, who started and stopped smoking, decided to quit. The conclusion is that more educated people, when they quit smoking, decide to do so at a younger age.

hypothesis. Using the data set constructed for this paper, the results obtained by Farrell and Fuchs (1986) can be replicated and confirmed.[14] However, this section will argue that the findings of Farrel and Fuchs (1986) are not sufficient to reject the causality from schooling to smoking because the decision to quit smoking is also strongly influenced by educational attainment and generally occurs when schooling is definitive.

Figure 6 illustrates that it is between age 30 and 35—between 1945 and 1955, when the information about the dangers of smoking started to diffuse—that the difference in smoking prevalence between college graduates and other individuals born between 1915 and 1920 started to emerge. This suggests that the decision to stop smoking—and the timing of this decision—is very important for explaining the differences in smoking prevalence across education categories and motivates a more formal analysis of the incidence of smoking cessation which combines the probability to quit with the age at quitting.

Following Meyer (1990) and Finkelstein and Poterba (2004), I use a semi-parametric proportional hazard model using Cox's partial likelihood technique (Cox 1972, 1975) to perform a survival analysis. Using the same data set, described previously, of smoking histories derived from subsequent National Health Interview Surveys, the sample is restricted to individuals who had smoked in their life and were still smoking at age 25, who are then observed from age 25 onward when their schooling level can be considered definitive.[15] The duration measure is the length of time between entry under observation at age 25 and exit for individuals who continued smoking until the date of the survey or the date of smoking cessation for former smokers at the date of interview. The failure event is smoking cessation.[16]

The hazard for individual i at time t is assumed to be:

$$(1) \qquad h_i(t) = h_o(t) \exp(x_i'\beta)$$

where $h_o(t)$ is the baseline hazard at time t, which is left unestimated, x_i is the vector of explanatory variables for individual i and β is the vector of parameters to be estimated. The implicit assumption behind this proportional hazard model is that the baseline hazard has the same shape for everybody and that covariates like education, gender, race, and age would affect it proportionally.

It is possible, however, to allow for the baseline hazard to take a different shape for different groups of individuals, by stratifying the estimation. In that case, the hazard for individual i at time t takes the form:

---

[14] Results available on request.

[15] Ninety-three percent of individuals who have smoked in their life started at age 25 or younger.

[16] I abstract from temporary interruptions in smoking that are shorter than one year (see note 5).

(2) $h_i(t) = h_{o1}(t) \exp(x_i'\beta)$ if individual *i* is in group 1

$h_i(t) = h_{o2}(t) \exp(x_i'\beta)$ if individual *i* is in group 2.

Table 2 presents results where the data have been stratified by gender, race, and year of birth as well as by age at interview and by the year of the interview, since the data is constructed from smoking histories taken from interviews ranging from 1978 to 2000. Table 2 indicates that, among smokers, the probability of smoking cessation at any point in time increases highly with education. The very significant results indicates that, for all decades and all birth cohorts, an additional year of education increases by 7.6 percent the probability that that person will quit smoking.[17]

Table 2 further analyzes the effect of education on the incidence of smoking cessation by dividing the sample across birth cohorts and across decades. The first row shows the effect of education on the probability of smoking cessation during the 1940-2000 period, by 10-year birth cohort, from the cohort born between 1910 and 1920 and the cohort born between 1970 and1980.[18] The effect of education appears to be increasing from earlier to later birth cohorts, with some leveling off for the three last birth cohorts. However, across birth cohorts the period considered as well as the age ranges considered varies widely. One constant, if we exclude the 1940-1950 decade, is that, reading each row horizontally from right to left, during each decade, the younger the individuals, the stronger the effect of education on smoking cessation.

The first column of table 2 shows the effect of education on the probability of smoking cessation for all birth cohorts, decade by decade, during the 1940-2000 period. The effect of education increases steeply from the 1940s to the 1950s and further from the 1950s to the 1960s, while the information on the dangers of smoking start to diffuse. The effect of education is still strong, but reaches a plateau thereafter. Again, however, the composition of the sample for each decade varies widely in terms of birth cohort and age ranges. In this respect, an analysis following each cohort separately from decade to decade is more interesting.

When table 2 is read vertically by birth cohort the striking result is for the four birth cohorts of smoking age during the 1950-1970 period. When the information on the health

---

[17] When looking at education categories, compared to individuals with less than a high school education, high school graduates are 23.1 percent are more likely to quit smoking; individuals with some college education are 46.8 percent more likely to quit smoking; and college graduates 106.1 percent more likely to stop smoking. These results also hold when the sample is divided across genders: the point estimates are slightly higher for females, but qualitatively they are similar.

[18] Notice that for this birth cohort, since I only consider individuals above age 25 and since the last survey used is from 2000, the survival analysis actually follows individuals born between 1970 and 1975 interviewed during the 1995-2000 surveys, while they were between ages 25 and 30. *Mutatis mutandis*, the same reasoning applies to the last cell of each row.

impact of smoking was not uniform and widespread, the effect of education on smoking cessation is always the highest during these two decades (see the highlighted box in table 2), even though during that period each birth cohort was in different age ranges. It seems therefore that, there is a specific period effect attached to the decades of the gradual diffusion of the information on the dangers of smoking. This is consistent with the hypothesis that a differential in the access to and in the ability to process information (Grossman 1972) might explain the positive correlation between health and education. This hypothesis will be further tested in section 4 that proposes an instrumental variable approach.

4. The Vietnam Draft as Instrument for College Education

The evidence presented in sections 2 and 3 documents the appearance, after 1950, of a negative gradient between educational achievement and smoking habits. This gradient is present at both margins: smoking initiation and smoking cessation. As such, this evidence is not entirely conclusive about the mechanisms through which education affect health decisions and does distinguish between three theories explaining the correlation between health and education: (1) education is causal in reducing smoking by entering as a factor in the health production function; (2) education increases the stream of future income and consumption and ultimately the value of life, so that it provides incentives for individuals to protect this stream by reducing mortality risks; or (3) the correlation between education and health is due to unobservables like the discount factor or ability that causes the same individuals to both study longer and take care of their health.

This section presents an instrumental variable approach using the Vietnam draft as an instrument for college education in order to test whether the negative relation between education and smoking is causal.

4.1 Ordinary Least Square Results

Table 3 reports results from ordinary least square regressions where the dependent variable is either an indicator for whether the individual is currently smoking (odd columns), or whether, among those who ever smoked, he did quit smoking (even columns). The data are taken from nine National Health Interview Surveys.[19] In order to allow the comparability with the results of the instrumental variable approach presented later, only individuals born between 1937 and 1956 are considered. The regressions use a simple linear probability model.

[19] The nine national surveys are from 1983, 1985, 1987, 1988, 1990, 1991, 1992, 1994 and 1995. The 1978, 1979 and 1980 surveys would have included individuals under age 25 for whom the educational level might not have been definitive. The public file of the 1997, 1998, 1999 and 2000 surveys did not contain the veteran status of the individuals, a variable that is crucial for the implementation of the instrumental variable approach. Notice that the "survivor bias" described in section 2 is only marginally a concern with this sub-sample of the data since individuals born between 1937 and 1956 would be aged between 39 and 58 in 1995.

Because, as documented in figure 2, it is mainly at the college margin that education seems to have an effect in decreasing smoking prevalence, the measure of education used in columns 1 to 4 of table 3 is the number of years of education above high school, so that individuals with no high school degree or just an high school degree are merged in the same category. In columns 7 and 8, the measure of education is an indicator for whether or not the individual is a college graduate or more. Columns 1 and 2 control for gender, race, family income, age at interview, birth cohort, and the date of the interview. The variable "trendmale" is included to account for the fact that over time, the positive difference in smoking prevalence between males and females has been decreasing (De Walque 2003). The measure of income is the logarithm of family income.[20] The results suggest that one year of college education decreases smoking prevalence by 4.0 percentage points and increases the probability of smoking cessation by 4.1 percentage points. These estimates, which are strongly significant, are consistent in magnitude with raw mean comparisons (see figure 2).

Columns 3 and 4 further control for the Vietnam veteran status of the individuals, but without controlling for income: while the coefficient on education remain in the same range, the results indicate that being a Vietnam veteran increases by 6.1 percentage points the probability of being a smoker, even after controlling for education, gender, race, age, birth cohort, and year of interview. This is a strong effect and will be discussed further, together with the instrumental variable approach. However, being a Vietnam veteran does not affect the probability of smoking cessation.

Columns 5 and 6 include controls for both income and actual Vietnam veteran status. The results indicate that, while income has a very significant negative effect on smoking prevalence and a positive one on the likelihood to have stopped smoking[21], the negative effect of education on smoking is very robust to the inclusion of income as a control. There is, in the education variable, something more than the ability to generate a higher income, which explains the smoking decision.

Columns 7 and 8 uses the indicator for college graduate or more as an alternative measure of education, and yields consistent results, since the coefficient on the dummy for college graduate is roughly four times larger than the coefficient on one year of college education, both for smoking prevalence and the probability of quitting.

---

[20] The National Health Interview Survey reports family income in $1,000 brackets between $1,000 and $20,000 and in $5,000 brackets between $20,000 and $50,000. The variable is top-coded at $50,000. Since the data comes from surveys ranging from 1983 to 1995, all reported incomes have been transformed in 1983 dollars using the inflation calculator of the Bureau of Labor Statistics.

[21] Introducing the measure of current income in a regression of smoking cessation where, typically, the decision to quit smoking has been taken in the past, might seem peculiar, unless current income is considered as a proxy for permanent income. To maintain the parallelism with the regressions on current smoking behavior, this specification is still presented (columns 2 and 6), but a specification without income as a control (column 4) is also available.

The main conclusion from these ordinary least square regressions is that the effect of college education in reducing smoking prevalence and increasing the probability of smoking cessation is robust even after controlling for gender, race, age, birth cohort, date of interview, and income. However, these results, if they allow us to conclude that the effect of education goes over and above the ability to generate an higher income, do not allow us to reject the hypothesis advanced by Fuchs (1982) that the fact that more educated individuals are less likely to smoke is due to selection on unobservables, like the discount factor.

Evidence from a plausibly exogenous variation in education would strengthen the argument that the effect of education in reducing smoking prevalence is causal. This motivates the implementation of an instrumental variable approach that exploits the fact that college attendance has been used as a draft avoidance strategy during the Vietnam War.[22]

4.2 An Instrumental Variable Approach: College Attendance as a Draft Avoidance Strategy

During the Vietnam War, some males enrolled in college to avoid the draft. Card and Lemieux (2000; 2001), by contrasting enrollment for males and females and the deviations from the trend in male/female ratio, estimate that this effect increased college graduation rates for males by 2 percentage points. They construct a measure of the induction risk faced by each birth cohort of males: the average yearly number of inductions in Vietnam during the years when a particular birth cohort was aged 19-22 divided by the size of the birth cohort. Based on their paper, I have reconstructed this measure.[23] It is plotted in figures 7 and 8. Of course, for females, the induction risk is always zero.

This measure of the risk of induction is used as an instrument for the effect of college education on smoking prevalence in table 4, which contains the results from the first stage and instrumental variable regressions.[24] The format is the same as in table 3.

The first stage regressions in the upper part of table 4, indicate that the risk of induction is always positively and significantly correlated with college education,

---

[22] In parallel research, Grimard and Parent (2003) develop a comparable strategy using a different data set. In their paper, which focuses exclusively on the strategy that uses the Vietnam draft as an instrument for college education, they define their instrument less precisely and obtain the result that education matters for smoking prevalence but not for smoking cessation. The latter result is somewhat surprising if one considers that, as argued in section 3, smoking cessation is the "cleaner" margin to look at the link between education and smoking, as the quitting decision is taken when schooling levels are definitive. They also exclude actual veterans from their sample in their main specifications.

[23] The number of inductions per semester is from appendix table 1 in Card and Lemieux (2000). The size of the cohort is taken from United States Government (2000).

[24] Notice that this instrument is different than the Vietnam-Era Draft Lottery used by Angrist and Krueger (1992). The draft lottery was in effect between 1970 and 1973, when the number of inductions was already declining.

indicating that males in birth cohort that experienced a higher exposure to the risk of being drafted in the army during the Vietnam War were more likely to remain longer in college. Graphically, the effect of the draft is illustrated in figure 7 in which the measure of the risk of induction is plotted together with males and females college graduation rates, computed from the sample. The highest level of the induction risk corresponds with the cohorts where the male college graduation rate is peaking, both in absolute terms and relative to the female college graduation rate. The instrument is in effect defined at the birth cohort and gender level. In addition to the effect of the exposure to the Vietnam draft, the instrument could thus pick birth-cohort-specific effects or secular trends in differences between genders. In order to avoid this, all regressions include dummies for each birth cohort, a dummy for male and a trend accounting for the secular changes in males/females ratios. The robust standard errors are also clustered by birth cohort and gender, as suggested by Moulton (1990).

The instrumental variable results are displayed in the lower part of table 4. In columns 1 and 2, no control for Vietnam veteran status are included. The results of this instrumental variable regression (positive and significant effect of college education on smoking, but not significant on smoking cessation) are different than the ordinary least square results (college education decreases significantly smoking prevalence and increases significantly smoking cessation). However, the risk of induction in the army, which is determined by the birth cohort, could also affect smoking directly and not only via education. In this case, the risk of induction would not be a valid instrument.

One potential channel through which the risk of induction can be directly correlated with smoking behavior is that the risk of induction is directly correlated with the probability of becoming a Vietnam veteran. The ordinary least square results reported in table 3 all show a strong positive effect of being a Vietnam veteran on smoking prevalence (but not on smoking cessation).[25] Without controlling for actually being a veteran, the instrument of the induction risk picks both the effect of attending college (decreases smoking) and the probability of being a veteran (increases smoking).

The National Health Interview Surveys, however, include the information on the actual veteran status of the individual, so the direct effect of being a Vietnam veteran on smoking behavior can be isolated. Columns 3-8 of table 4 therefore control for actual Vietnam veteran status. Columns 3 and 4 do not control for income. In column 3, the instrumental variable estimate of the effect of college education on smoking prevalence is negative and significant, and is somewhat larger than the ordinary least square estimate (–0.06 versus –0.047). The significance is lower in column 4, when smoking cessation is analyzed. When controlling for family income, as in columns 5 and 6, the instrumental variable estimates are very close to the ordinary least square estimates and the actual mean

[25] Bedard and Deschenes (2003) also show that World War II veterans are more likely to smoke.

differences in the raw data, but the result for smoking cessation is still only borderline significant. Comparable results are obtained in columns 7-8 with an indicator for college graduation as a measure of education.

Table 5 presents results using an alternative instrument: rather than only the risk of being drafted, I have created an instrument that also includes the risk of being killed in Vietnam. The intuition behind this alternative instrument is that what might have caused individuals to enroll in college to avoid the draft is not so much the risk of being drafted, but the risk of being drafted and killed. I calculate the risk of being killed in action as the ratio between the number of soldiers killed in action in a year and the level of troops engaged in Vietnam during that year.[26] Then, for each year, I multiply the yearly induction rate with the probability to be killed in action.[27]

Figure 8 compares the two risk measures used as instrument, the risk of induction and the risk of induction multiplied by the risk of being killed in action. Although these measures peak for the same birth cohort, the slope is steeper when the probability of being killed in action is also considered. As reported in table 5, when the risk of induction multiplied by the risk of being killed in action is used as an instrument (see columns 3 to 8, when actual Vietnam veteran status is controlled for), the results are qualitatively very similar to the results with the induction risk (see table 4) as instrument. However, the results are here always significant, even for the smoking cessation regressions. If anything, the point estimates of the coefficients are slightly higher than the ordinary least square results and the first instrumental variable strategy, especially in the case of smoking cessation. It is especially important to establish the significance of the positive effect of college education on the decision to stop smoking, because, as argued in section 3, this is the more appropriate margin to study the relationship between education and smoking, since that decision is taken after schooling is completed. Interestingly, this is also a margin where the actual Vietnam veteran status has no effect, as indicated by the results in the even columns in tables 3 to 5.

On the other hand, actual Vietnam veteran status has an important impact of the probability of being a current smoker, even after controlling for age, income, education, race, and gender. The inclusion of this variable in the instrumental variable regressions

---

[26] The casualty data are taken from the website http://www.vietnamwall.org/casualty.html (accessed on May 11, 2004) and derived from the Combat Area Casualty File of 11/93 and the Adjutant General's Center (TAGCEN) file of 1981, available from the National Archives. For the years 1961-1965, the number of killed in action (1864) has been evenly spread over the five years, as no yearly details were available. The level of U.S. troops engaged in Vietnam for each year is taken from the website http://www.25thaviation.org/id430.htm (accessed on May 11, 2004).

[27] The final measure of the risk to be induced and killed for each birth cohort is calculated as the previous measure of the risk of induction. I average the yearly number of inductions in Vietnam multiplied by the probability of being killed in action over the years when a particular birth cohort was aged 19-22, and divide this number by the size of the birth cohort.

appears to be crucial for its success as indicated by the inconsistent results obtained when, as in columns 1 and 2, Vietnam veteran status is not controlled for.[28]

4.3 Interpretation of the Instrumental Variables Estimates of the Effect of College Education on Smoking

The instrumental variable strategy implemented in this chapter uses the fact that college enrollment as a draft avoidance strategy during the Vietnam War era provides a plausibly exogenous variation in educational attainment. During the Vietnam War, going to college was associated with an additional benefit: avoiding the draft and the disutility associated with it. This additional benefit induced some individuals that would otherwise not have enrolled in college to enroll.

It is however likely that the individuals induced to enroll in college because of the Vietnam draft were not random among the group of individuals who, in the absence of the draft, would not have gone to college. On the contrary, it is reasonable to expect that those males who decided to enroll because of the draft were the ones who were close (just under) to the margin of going to college. The possibility to avoid the draft if they were in college made them switch and decide to enroll in college. Typically, the "switchers" would have a lower discount rate (i.e., a larger β) than the individuals who did not enroll, despite the extra benefit of draft avoidance, but a higher discount rate than the ones who would have gone to college even in the absence of the draft.

It then follows that, if there is heterogeneity in the returns to schooling, the instrumental variable estimate obtained is not a consistent estimate of the average effect of education on smoking in the population (Card 1995, 2001; Heckman, Lalonde, and Smith 1999). The instrumental variable estimate should rather be interpreted as the effect of education for the subgroup of males that has been induced to enroll in college in order to avoid the Vietnam Draft. Imbens and Angrist (1994) define this estimator as the local average treatment effect (LATE).

This qualification in the interpretation of the results is important, as it does not permit generalizing the estimates obtained from the instrumental variable estimates for the entire population. However, an estimation of the effect of education on smoking for those individuals who are at the margin between going to college or not is interesting from a

[28] However, since the first stage of the instrumental variable strategy implies that it was possible to avoid, or at least to delay, the draft by staying in college, it might be argued that Vietnam veteran status is endogenous. To address the issue of the potential endogeneity of veteran status, I would need an additional instrument for veteran status. I am investigating the feasibility to use height and congenital medical conditions available in some of the National Health Interview Surveys, as those physical characteristics, associated with a higher probability of failing the physical induction test in the army, might be an exogenous determinant of the likelihood to be a Vietnam veteran.

policy perspective, as this is the group that is likely to be induced to enroll by other interventions such as tuition subsidies.

The results of the instrumental variable approach implemented in this section suggest that at least some of the effect of college education is causal in reducing smoking prevalence. A similar result is obtained by Currie and Moretti (2003) who investigate smoking during pregnancy using the dates of college openings as an instrument for maternal education. Both instruments are likely to be effective at the college margin, which is the relevant margin for smoking decisions as documented in figure 2.

5. Conclusions

This paper starts by analyzing smoking histories reconstructed from National Health Interview Surveys in the United States and retracing the evolution of smoking patterns across education groups. The main conclusion is that, after the gradual arrival of the information about the dangers of tobacco, the smoking prevalence among more educated people declined earlier and more dramatically.

The time pattern of the decline in smoking prevalence for individuals with college degrees suggests that at least some of the differential by education can be ascribed to a causal role of education in giving access to and processing the information. Figure 2 shows that the smoking prevalence among college graduates started to decline earlier than for the three other categories. The inflection point is in 1954, 10 years before the first Surgeon General's Report, four years after the consensus in the medical profession, and one year after the first article in the general press on the dangers of smoking. Further, the results of multivariate regressions show that the effect of college education on smoking prevalence is robust even after controlling for income, indicating that the effect of education goes over and above the ability to generate a higher income.

This paper also argues that, because the decision to quit smoking is usually taken after education is completed, this margin of the smoking decision is cleaner to analyze than smoking initiation that often occurs before schooling levels are definitive. A detailed study of the incidence of smoking cessation using a proportional hazard model shows a very strong effect of education on the probability, for smokers, to quit smoking at any point in time. Very interestingly, for all birth cohorts that were of smoking age during the 1950-1970 period, it is during this interval that the effect of education is the strongest, independent of age, suggesting that the impact of education was at its peak when access to information about the risks associated with tobacco was not uniform.

Finally, in order to isolate an exogenous variation in college education, this paper exploits the fact that college attendance was used as a draft avoidance strategy during the Vietnam War period. Using the risk of induction in Vietnam as an instrument for college education, and controlling for actual Vietnam veteran status, instrumental variable

estimates of the effect of education on smoking behavior are very similar to ordinary least square estimates, suggesting that the effect of college education in reducing smoking prevalence is causal.

Rates of return to education, as traditionally calculated, only account for labor market earnings. If the effect of education on health and longevity is, as in the case of smoking reduction, causal, it makes sense to attribute additional returns to education. I have proposed (De Walque 2003; 2004) a method to estimate these extra returns in the context of the HIV/AIDS epidemic in Africa. Compared to HIV/AIDS case, the potential addition to the returns to education is significant but modest in the case of smoking, essentially because smoking kills at relatively late ages.

Bibliography

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Angrist, Joshua D. 1990. "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records." *American Economic Review* 80(3): 313-36.

Angrist, Joshua D., and Krueger, Alan B. 1992. "Estimating the Payoff to Schooling Using the Vietnam-Era Draft Lottery." NBER Working Paper 4067. Cambridge, Mass.

Becker, Gary S. 1993. *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*. 3rd Edition. Chicago, IL: The University of Chicago Press.

Becker, Gary S., Grossman Michael, and Murphy Kevin M. 1994. "An Empirical Analysis of Cigarette Addiction." *American Economic Review* 84(3): 396-418.

Bedard, Kelly, and Olivier Deschenes. 2003. "The Long-Term Impact of Military Service on Health: Evidence from World War II Veterans." Processed.

Berger, Mark C., and J. Paul Leigh. 1989. "Schooling, Self-Selection and Health." *The Journal of Human Resources* 24(3): 433-55.

Card, David. 1995. "Earnings, Schooling, and Ability Revisited." *Research in Labor Economics* 14: 23-48

Card, David. 2001. "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems." *Econometrica* 69(5): 1127-60.

Card, David, and Lemieux, Thomas. 2000. "Going to College to Avoid the Draft: The Unintended Legacy of the Vietnam War." Processed.

Card, David, and Thomas Lemieux. 2001. "Draft Avoidance and College Attendance: The Unintended Legacy of the Vietnam War." *American Economic Review Papers and Proceedings* 91: 97-102.

Chaloupka, F. J., and K. E. Warner. 2000. "The Economics of Smoking." In J. Newhouse and A. Culyer, eds., *Handbook of Health Economics 1A*. North-Holland, Amsterdam..

Cleves, Mario, William W. Gould, and Roberto G. Gutierrez. 2002. *An Introduction to Survival Analysis Using Stata*. College Station, TX: Stata Press Publication.

Currie, Janet, and Enrico Moretti. 2003. "Mother's Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings." *Quarterly Journal of Economics* 118(4): 1495-1532.

Cutler, David M., and Srikanth Kadiyala. 2003. "The Return to Biomedical Research: Treatment and Behavioral Effects." In Kevin Murphy and Robert Topel, eds., *Measuring the Gains from Medical Research. An Economic Approach.* Chicago, IL: University of Chicago Press.

Cox, D. R. 1972. "Regression Models and Life-Tables." *Journal of the Royal Statistical Society B* 34: 187-202.

Cox, D. R. 1975. "Partial Likelihood." *Biometrika* 62: 269-76.

Deaton, Angus, and Christina Paxson. 2003. "Mortality, Income, and Income Inequality Over Time in Britain and the United States." In David A. Wise, ed., *Perspectives on The Economics of Aging,* vol. 8. Chicago, IL: University of Chicago Press for NBER.

De Walque, Damien. 2003. "How Do Education and Information Affect Health Decisions? The Cases of HIV/AIDS and Smoking." Ph.D. Dissertation, University of Chicago, Department of Economics.

De Walque, Damien. 2004. "How Does the Impact of an HIV/AIDS Information Campaign Vary with Educational Attainment. Evidence from Rural Uganda." Policy Research Working Paper 3289. World Bank, Washington DC.

Farell, Phillip, and Victor Fuchs. 1986. "Schooling and Health: The Cigarette Connection." In Victor R. Fuchs, ed., *The Health Economy.* Cambridge, Mass. and London: Harvard University Press.

Ferrence, Roberta G. 1989. *The Rise and Fall of Cigarette Smoking in North America.* New York, London: Garland Publishing.

Finkelstein, Amy, and James Poterba. 2004. "Adverse Selection in Insurance Markets: Policyholder evidence from the U.K. Annuity Market." *The Journal of Political Economy* 112 (1,1): 183-208.

Fuchs, Victor. 1982. "Time Preference and Health: An Exploratory Study." In Victor Fuchs, ed., *Economic Aspects of Health*. Chicago: University of Chicago Press.

Grimard, Franque, and Daniel Parent. 2003. "Education and Smoking: Were Vietnam Draft Avoiders Also More Likely to Avoid Smoking?" Scientific Series 2003s-44, Centre interuniversitaire de recherché en analyse des organisations (Cirano), Montréal.

Grossman, Michael. 1972. "On the Concept of Health Capital and the Demand for Health." *The Journal of Political Economy* 80(2): 223-255.

Grossman, Michael. 1975. "The Correlation between Health and Smoking." In Nestor E. Terlecky, ed., *Household Production and Consumption.* New York: National Bureau of Economic Research.

Grossman, Michael. 2004. "Education and Personal Behavior." In E. Hanushek and F. Welch, eds., *Handbook of the Economics of Education*. Forthcoming Amsterdam: Elsevier Science.

Harris, J. E. 1983. "Cigarette smoking among successive birth cohorts of men and women in the United States during 1900-80." *Journal of the National Cancer Institute* 71: 473-479.

Heckman, James J., Robert J. Lalonde, and Jeffrey A. Smith. 1999. "The Economics and Econometrics of Active Labor Market Programs." In Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, volume III. Amsterdam: Elsevier Science.

Imbens, Guido and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62: 467-76.

Kenkel, Donald S. 1991. "Health Behavior, Health Knowledge, and Schooling." *The Journal of Political Economy* 99(2): 287-305.

Kiefer, Nicholas M. 1988. "Economic Duration Data and Hazard Functions." *Journal of Economic Literature* 26(2): 646-79.

Lleras-Muney, Adriana. 2004. "The Relationship Between Education and Mortality in the U.S." *Review of Economic Studies,* forthcoming.

Meara, Ellen. 2001. "Why is Health Related to Socioeconomic Status? The Case of Pregnancy and Low Birth Weight." NBER Working Paper 8231. Cambridge, Mass.

Meyer, Bruce D. 1990. "Unemployment Insurance and Unemployment Spells." *Econometrica* 58(4): 757-82.

Moulton, Brent R. 1990. "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Unit." *The Review of Economics & Statistics* 72(1): 334-38.

National Conference on Smoking & Health. 1970. *A Summary of Proceedings*. Sept. 9-11, 1970. Washington, D.C.

Peto, Richard, Sarah Darby, Harz Deo, Paul Silcoks, Elise Whitley, and Richard Doll. 2000. "Smoking, smoking cessation, and lung cancer in the UK since 1950: combination of national statistics with two case-control studies." *British Medical Journal* 321: 323-29.

Sander, William. 1995. "Schooling and Quitting Smoking." *The Review of Economics and Statistics* 77(1): 191-199.

United States Government. 1964. *Smoking and Health. Report of the Advisory Committee to the Surgeon General of the Public Health Service.* Public Health Service Publication No. 1103. Department of Health, Education, and Welfare, Public Health Service. Washington, D.C. : UGSPO.

United States Government. 1980. *The Health Consequences of Smoking for Women: A Report of the Surgeon General*. Department of Health and Human Services. Rockville, Maryland.

United States Government. 1989. *Reducing the Health Consequences of Smoking: 25 Years of Progress, a Report of the Surgeon General*. Department of Health and Human Services Publication No. (CDC) 89-8411.

United States Government. 1994. *CDC Surveillance Summaries, Surveillance for Selected Tobacco-Use Behaviors - United States, 1900-1994*, Center for Disease Control and Prevention, MMWR 43 (No. SS-3).

United States Government. 2000. *Digest of Education Statistics*. Department of Education, Washington D.C.: UGSPO.

United States Government. 2004. *History of the 1964 Surgeon General's Report on Smoking and Health*. Center for Disease Control and Prevention (http://www.cdc.gov/tobacco/30yrsgen.htm, accessed on May 11, 2004)

United States Government. *National Health Interview Survey, 1978, 1979, 1980, 1983, 1985, 1987, 1988, 1990, 1991, 1992, 1994, 1995, 1997, 1998, 1999, 2000*, Department of Health and Human Services, National Center for Health Statistics. Hyattsville, MD.

Viscusi, W. K. 1990. "Do Smokers Underestimate Risks?" *The Journal of Political Economy* 98(6): 1253-1269.

Viscusi, W. K. 1992. *Smoking, Making the Risky Decision*. Oxford University Press.

Wald, Nicholas, Stephanie Kiryluk, Sarah Darby, Richard Doll, Malcolm Pike, and Richard Peto. 1988. *UK Smoking Statistics*. Oxford University Press.

Table 1. The Impact of Education on Smoking Initiation and Smoking Cessation

Cohort born in	All	1910-19	1920-29	1930-39	1940-49	1950-59	1960-69
Age in 1950		31-40	21-30	11 to 20	0 -11	n.a.	n.a.
Age in 1970		51-60	41-50	31-40	21-30	11 to 20	0 -11
Dependent variable: Ever started smoking							
High school	–0.028***	0.006	-0.018***	-0.037***	-0.037***	-0.047***	-0.057***
	[–12.63]	[1.09]	[-3.59]	[-6.90]	[-6.73]	[-9.40]	[-9.57]
Some college	–0.043***	0.042***	0.007***	-0.01	-0.044***	-0.091***	-0.126***
	[–17.04]	[5.32]	[1.09]	[-1.49]	[-7.41]	[-17.42]	[-21.76]
College and above	–0.178***	-0.027***	-0.064***	-0.112***	-0.17***	-0.253***	-0.295***
	[–72.58]	[-3.20]	[-9.09]	[-16.63]	[-29.01]	[-51.10]	[-46.35]
Male	0.177***	0.336***	0.296***	0.221***	0.173***	0.103***	0.036***
	[107.14]	[62.75]	[67.26]	[50.05]	[46.68]	[31.28]	[8.75]
Black	–0.044***	-0.06***	-0.034***	-0.039***	-0.034***	-0.032***	-0.1***
	[–18.14]	[-7.57]	[-5.02]	[-5.95]	[-6.13]	[-6.82]	[-18.03]
Other race	–0.122***	-0.097***	-0.118***	-0.167***	-0.148***	-0.112***	-0.111***
	[–28.20]	[-4.77]	[-8.07]	[-13.31]	[-14.88]	[-13.96]	[-13.27]
Pseudo R-square	0.04	0.09	0.07	0.04	0.04	0.03	0.05
Observations	348,198	35,991	50,255	49,774	68,778	89,381	54,019
Observed probability	0.52	0.472	0.561	0.586	0.57	0.502	0.421
Dependent variable: Stopped smoking (for those who ever started)							
High school	0.068***	0.024***	0.049***	0.091***	0.097***	0.097***	0.091***
	[24.34]	[3.99]	[8.47]	[13.90]	[14.19]	[15.19]	[8.40]
Some college	0.128***	0.041***	0.073***	0.136***	0.159***	0.178***	0.164***
	[40.45]	[5.03]	[9.87]	[17.52]	[21.69]	[25.00]	[16.38]
College and above	0.244***	0.075***	0.136***	0.222***	0.295***	0.328***	0.298***
	[75.32]	[8.76]	[18.49]	[29.75]	[42.92]	[43.65]	[24.68]
Male	0.05***	0.105***	0.118***	0.091***	0.032***	-0.004***	-0.011**
	[22.89]	[16.55]	[22.09]	[16.45]	[6.75]	[-1.14]	[-1.96]
Black	–0.095***	-0.072***	-0.072***	-0.075***	-0.104***	-0.104***	-0.104***
	[–28.63]	[-6.46]	[-7.96]	[-8.68]	[-13.99]	[-18.31]	[-12.28]
Other race	–0.03***	-0.007	-0.02	-0.034*	-0.036**	-0.023**	-0.015
	[–4.49]	[-0.27]	[-0.94]	[-1.73]	[-2.42]	[-2.08]	[-1.18]
Pseudo R-square	0.09	0.07	0.07	0.06	0.05	0.05	0.04
Observations	181,222	16,988	28,212	29,177	39,206	44,884	22,755
Observed probability	0.464	0.677	0.599	0.514	0.442	0.365	0.308

***, **, * denote statistical significance at the 1%, 5% and 10% confidence levels, respectively.

Note: Analysis from smoking histories from the 1978, 1979, 1980, 1983, 1985, 1987, 1988, 1990, 1991, 1992, 1994, 1995, 1997, 1998, 1999 and 2000 National Health Interview Surveys for individuals age 25 and above at the time of the survey. Average of marginal coefficients, observation by observation, from a logit estimation. Robust z-statistics in square brackets. Survey and year of birth dummies included. The excluded dummies are less than high school, female and white.

Table 2. Incidence of Smoking Cessation by Decade and Birth Cohort

		Birth cohort						
Decade	All	1910-20	1920-30	1930-40	1940-50	1950-60	1960-70	1970-80
1940-2000	1.076***	1.034***	1.05***	1.085***	1.127***	1.162***	1.163***	1.156***
Z-stat	[52.99]	[12.33]	[19.62]	[27.19]	[34.47]	[31.04]	[15.33]	[3.46]
N	162,993	18,158	28,825	29,394	37,674	39,282	18,115	2,290
Failures	63,741	11,774	16,175	13,548	13,213	8,887	2,437	112
Time at risk	3,174,642	621,820	862,255	690,859	617,511	400,659	112,211	5,558
1940-1950	1.036***	1.039***	1.003					
Z-stat	[4.58]	[3.79]	[0.15]					
N	32,942	18,158	12,555					
Failures	1,426	878	219					
Time at risk	216,140	146,815	36,655					
1950-1960	1.06***	1.053***	1.062***	1.094***				
Z-stat	[12.14]	[7.01]	[8.16]	[4.81]				
N	58,107	17,280	28,606	13,047				
Failures	4,362	1,743	2,033	3,49				
Time at risk	448,876	166,460	239,013	3,9155				
1960-1970	1.078***	1.057***	1.067***	1.121***	1.176***			
Z-stat	[24.49]	[10.18]	12.82	[18.84]	[10.92]			
N	82,989	15,537	26,573	29,045	15,755			
Failures	1,1335	3,042	4,127	3,384	726			
Time at risk	651,869	14,4175	250,288	237,646	45,401			
1970-1980	1.081***	1.027***	1.053***	1.092***	1.147***	1.184***		
Z-stat	[31.24]	[5.44]	[11.36]	[16.75]	[26.18]	[13.98]		
N	108,763	12,495	22,446	25,661	36,948	18,913		
Failures	18,954	3,561	4,780	4,547	5,640	1,280		
Time at risk	834,046	108,368	202,809	235,695	289,536	53,829		
1980-1990	1.08***	0.999	1.033***	1.063***	1.118***	1.167***	1.191***	
Z-stat	[29.54]	[-0.05]	[6.65]	[11.56]	[20.34]	[26.02]	[10.37]	
N	107296	7228	15001	18242	27595	36412	12005	
Failures	20698	2234	4100	4087	5291	5711	783	
Time at risk	740,146	48,770	110,048	140,578	217,681	251,814	34,669	
1990-2000	1.073***	1.015	1.02*	1.026**	1.057***	1.126***	1.15***	1.156***
Z-stat	[14.23]	[0.81]	[1.92]	[2.50]	[5.56]	[10.61]	[11.56]	[3.46]
N	54779	1740	4945	7324	11981	17317	14548	2290
Failures	6966	316	916	1181	1556	1896	1654	112
Time at risk	283,565	7,232	23,442	37,785	64,893	95,016	77,542	5,558

***, **, * denote statistical significance at the 1%, 5% and 10% confidence levels, respectively.

Note: Analysis from smoking histories from the 1978, 1979, 1980, 1983, 1985, 1987, 1988, 1990, 1991, 1992, 1994, 1995, 1997, 1998, 1999 and 2000 National Health Interview Surveys for individuals age 25 and above at the time of the survey. Smoking histories are reconstructed using the age at smoking initiation and cessation. See also note 5. Proportional hazard model (Cox model) stratified by gender, year of birth, race, age at interview and survey year. Based on individuals who were smoking at age 25. Failure is smoking cessation. Robust z-statistics in square brackets.

Table 3. Current Smoking and Smoking Cessation for Individuals Born between 1937 and 1956, Ordinary Least Square Results

	1	2	3	4	5	6	7	8
Dependent variable	smoking now	stopped smoking	smoking now	stopped smoking	smoking now	stopped smoking	smoking now	stopped smoking
Years of education above High school	–0.04***	0.041***	–0.047***	0.05***	–0.04***	0.041***	_	_
	[–10.53]	[19.35]	[–14.12]	[19.87]	[–11.08]	[20.11]		
College and above	–	–	–	–	–	–	–0.173***	0.178***
							[–11.68]	[18.56]
Male	0.102***	0.048***	0.073***	0.063***	0.079***	0.051***	0.077***	0.052***
	[14.56]	[10.56]	[12.08]	[14.16]	[11.16]	[9.79]	[10.89]	[9.59]
Trendmale	–0.001**	–0.003***	–0.0002	–0.003***	–0.0003	–0.003***	–0.0003	–0.003***
	[–2.14]	[–7.44]	[–0.53]	[–9.41]	[–0.84]	[–7.62]	[–0.69]	[–7.68]
Log family income in 1983$	–0.076***	0.084***	_	_	–0.077***	0.084***	–0.084***	0.091***
	[–27.94]	[25.10]	_	_	[–26.59]	[25.37]	[–22.52]	[28.31]
Vietnam veteran	_	_	0.061***	–0.004	0.063***	–0.005	0.059***	–0.001
	_	_	[5.83]	[–0.60]	[5.94]	[–0.73]	[5.59]	[–0.24]
Black	–0.002	–0.051***	0.036***	–0.095***	–0.0009	–0.051***	–0.0005	–0.052***
	[–0.32]	[–8.66]	[5.65]	[–16.69]	[–0.13]	[–8.75]	[–0.06]	[–9.05]
Other race	–0.07**	–0.014	–0.051	–0.039**	–0.066**	–0.015	–0.068**	–0.01
	[–2.63]	[–1.10]	[–1.60]	[–2.71]	[–2.34]	[–1.12]	[–2.54]	[–0.78]
R-square	0.068	0.09	0.055	0.074	0.069	0.09	0.065	0.08
Observations	74,810	40,753	83,069	45,108	73,952	40,319	73,957	40,323

***, **, * denote statistical significance at the 1%, 5% and 10% confidence levels, respectively.

_ indicates that the variable was not included in the regression.

Note: Analysis from smoking histories from the 1983, 1985, 1987, 1988, 1990, 1991, 1992, 1994, and 1995 National Health Interview Surveys for individuals age 25 and above at the time of the survey. Individuals born between 1937 and 1956. Linear regression model, where the dependent variable is current smoking behavior (odd columns) or smoking cessation for individuals who ever smoked (even columns). Robust t-statistics, clustered by birth cohort and gender, in square brackets. Birth cohort, age at interview and survey dummies included in all specifications. Trendmale, a dummy for male multiplied by a yearly trend where born in 1937=0, accounts for the trend in the difference in smoking behavior between males and females. See note 20 for details on the family income variable. Sample sizes vary across specifications as the Vietnam veteran status and the income variable might have been missing.

Table 4. Instrumental variable estimation: instrument for years of education is the risk of induction in Vietnam

	1	2	3	4	5	6	7	8
	First stage regression							
Dependent variable	Years of education above high school						College and above	
Induction risk	1.095***	0.742***	1.64***	0.988***	1.779***	1.227***	0.399***	0.284***
	[4.86]	[2.64]	[7.24]	[3.52]	[7.68]	[4.25]	[8.07]	[4.67]
R-square	0.116	0.095	0.032	0.023	0.118	0.096	0.092	0.073
Observations	74,810	40,753	83,069	45,108	73,952	40,319	73,957	40,323
Instrumental variable regression								
Dependent variable	smoking now	stopped smoking	smoking now	stopped smoking	smoking now	stopped smoking	smoking now	stopped smoking
Years of education above high school	0.076*	0.028	–0.06***	0.052	–0.038*	0.051	_	_
	[1.83]	[0.61]	[–3.10]	[1.38]	[–1.88]	[1.62]	_	_
College and above	_	_	_	_	_	_	–0.169*	0.219
	_	_	_	_	_	_	[–1.82]	[1.57]
Male	0.058***	0.051***	0.081***	0.063***	0.078***	0.049***	0.077***	0.05***
	[3.64]	[6.00]	[6.98]	[5.35]	[6.83]	[6.35]	[6.89]	[6.92]
Trendmale	0.00003	–0.003***	–0.0004	–0.003***	–0.0003	–0.003***	–0.0003	–0.003***
	[0.05]	[–7.80]	[–0.85]	[–8.94]	[–0.64]	[–7.64]	[–0.55]	[–8.06]
Log family income in 1983$	–0.178***	0.093**	_	_	–0.078***	0.077***	–0.085***	0.086***
	[–4.71]	[2.66]	_	_	[–4.14]	[3.22]	[–5.21]	[4.47]
Vietnam veteran	_	_	0.057***	–0.003	0.063***	–0.003	0.06***	0.001
	_	_	[7.16]	[–0.56]	[7.57]	[–0.49]	[6.97]	[0.12]
Black	0.024	–0.053***	0.027*	–0.094***	–0.0005	–0.05***	–0.0003	–0.05***
	[1.24]	[–6.44]	[1.73]	[–4.14]	[–0.06]	[–6.59]	[–0.03]	[–7.19]
Other race	–0.145***	–0.008	–0.044	–0.039*	–0.067**	–0.019	–0.069	–0.013
	[–4.85]	[–0.28]	[–1.41]	[–1.77]	[–2.29]	[–0.84]	[–2.45]	[–0.69]
R-square	(.)	0.088	0.052	0.074	0.068	0.089	0.065	0.084
Observations	74,810	40,753	83,069	45,108	73,952	40,319	73,957	40,323

***, **, * denote statistical significance at the 1%, 5% and 10% confidence levels, respectively.

(.) denotes negative R-square.

_ indicates that the variable was not included in the regression.

Note: Analysis from smoking histories from the 1983, 1985, 1987, 1988, 1990, 1991, 1992, 1994, and 1995 National Health Interview Surveys for individuals age 25 and above at the time of the survey. Individuals born between 1937 and 1956. Linear regression model, where the dependent variable is current smoking behavior (odd columns) or smoking cessation for individuals who ever smoked (even columns). Robust t-statistics, clustered by birth cohort and gender, in square brackets. Birth cohort, age at interview and survey dummies included in all specifications. Trendmale, a dummy for male multiplied by a yearly trend where born in 1937=0, accounts for the trend in the difference in smoking behavior between males and females. See note 20 for details on the family income variable. The instrument for education is the induction risk in Vietnam faced by each birth cohort of males: it is the average yearly number of inductions in Vietnam during the years when that birth cohort was aged 19-22, divided by the size of that birth cohort. See note 23 for the data source for the instrument. Sample sizes vary across specifications as the Vietnam veteran status and the income variable might have been missing.

Table 5. Instrumental variable estimation: instrument for years of education is the risk of induction in Vietnam multiplied by the risk of being killed in action

	1	2	3	4	5	6	7	8
	First stage regression							
Dependent variable	Years of education above high school						College and above	
Induction risk* risk of being killed in action	54.74*** [4.85]	40.69*** [2.87]	80.02*** [7.04]	51.22*** [3.62]	87.28*** [7.53]	64.66*** [4.45]	19.27*** [7.77]	14.51*** [4.73]
R-square	0.116	0.095	0.032	0.023	0.118	0.096	0.092	0.073
Observations	74,810	40,753	83,069	45,108	73,952	40,319	73,957	40,323
	Instrumental variable regression							
Dependent variable	smoking now	stopped smoking	smoking now	stopped smoking	smoking now	stopped smoking	smoking now	stopped smoking
Years of education above high school	0.069* [1.80]	0.055 [1.29]	-0.059*** [-3.11]	0.068* [1.94]	-0.04** [-2.11]	0.066** [2.24]	_ _	_ _
College and above	_ _	_ _	_ _	_ _	_ _	_ _	-0.181* [-2.07]	0.297** [2.24]
Male	0.061*** [3.91]	0.046*** [5.54]	0.081*** [6.98]	0.056*** [4.97]	0.079*** [7.25]	0.044*** [5.75]	0.078*** [7.23]	0.045*** [6.16]
Trendmale	-0.00004 [-0.08]	-0.003*** [-7.35]	-0.0004 [-0.85]	-0.003*** [-8.49]	-0.0003 [-0.71]	-0.003*** [-7.26]	-0.0003 [-0.62]	-0.003*** [-7.75]
Log family income in 1983$	-0.171*** [-4.96]	0.074** [2.31]	_ _	_ _	-0.077*** [-4.31]	0.077*** [2.90]	-0.083*** [-5.40]	0.076*** [4.16]
Vietnam veteran	_ _	_ _	0.058*** [7.28]	-0.001 [-0.16]	0.062*** [7.92]	-0.003 [-0.04]	0.059*** [7.43]	0.006 [0.73]
Black	0.022 [1.20]	-0.049*** [-5.32]	0.027* [1.76]	-0.084*** [-3.90]	-0.001 [-0.11]	-0.047*** [-5.71]	-0.0009 [-0.09]	-0.048*** [-6.29]
Other race	-0.14*** [-4.79]	-0.021 [-0.74]	-0.045 [-1.42]	-0.046 [-2.01]*	-0.066 [-2.25]**	-0.027 [-1.15]	-0.067 [-2.40]**	-0.02 [-1.00]
R-square	(.)	0.087	0.053	0.069	0.069	0.081	0.065	0.076
Observations	74,810	40,753	83,069	45,108	73,952	40,319	73,957	40,323

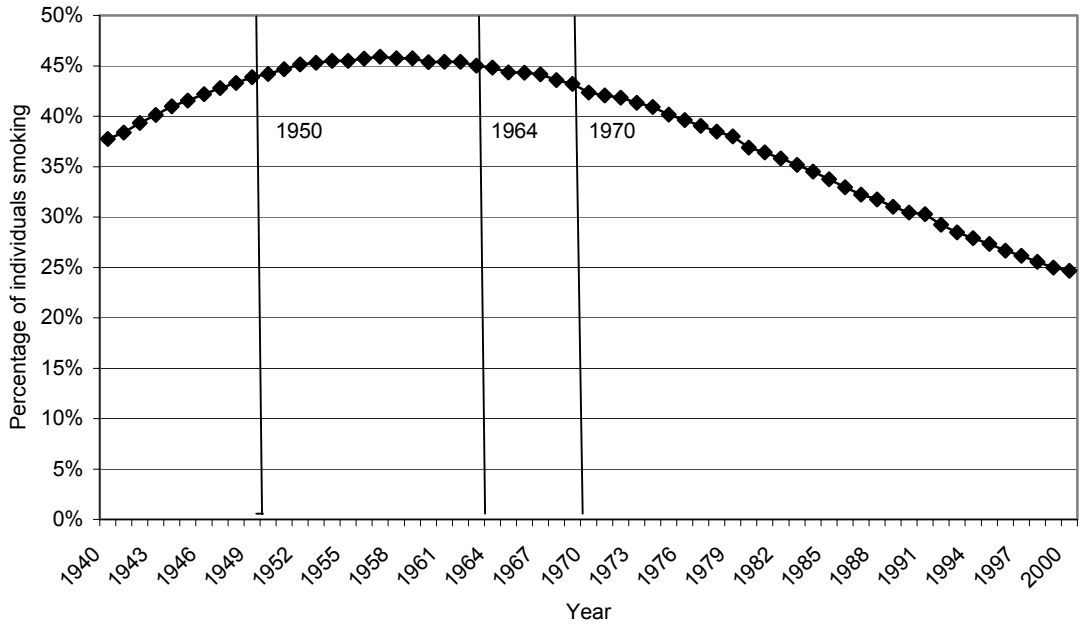
***, **, * denote statistical significance at the 1%, 5% and 10% confidence levels, respectively.

(.) denotes negative R-square.

_ indicates that the variable was not included in the regression.

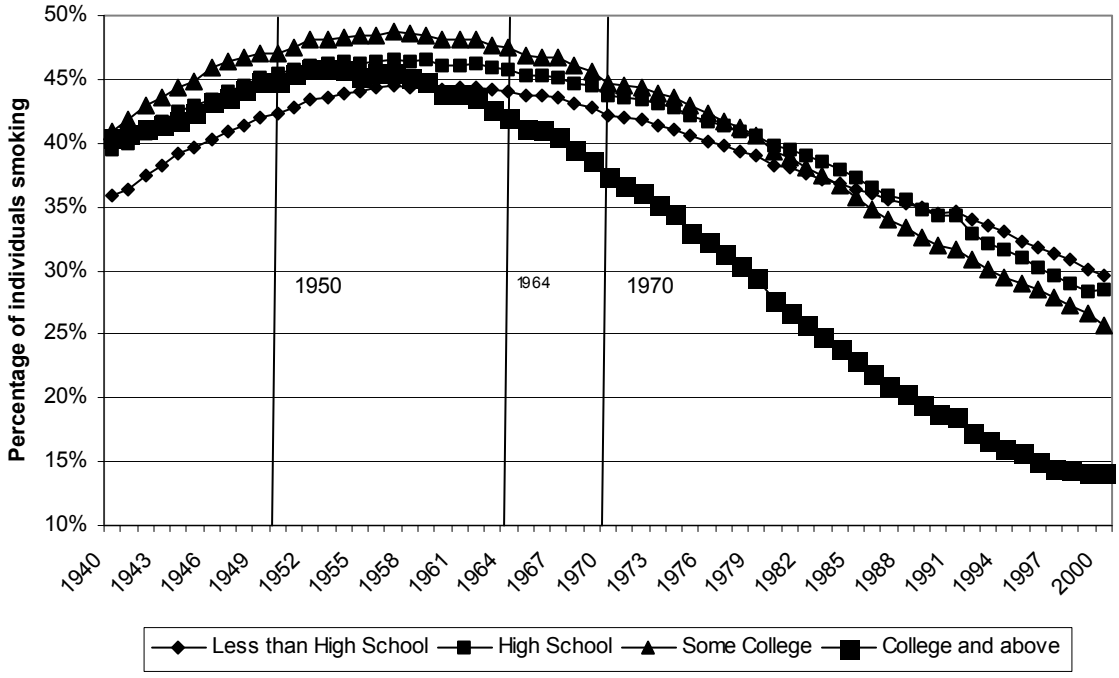
Note: Analysis from smoking histories from the 1983, 1985, 1987, 1988, 1990, 1991, 1992, 1994, and 1995 National Health Interview Surveys for individuals age 25 and above at the time of the survey. Individuals born between 1937 and 1956. Linear regression model, where the dependent variable is current smoking behavior (odd columns) or smoking cessation for individuals who ever smoked (even columns). Robust t-statistics, clustered by birth cohort and gender, in square brackets. Birth cohort, age at interview and survey dummies included in all specifications. Trendmale, a dummy for male multiplied by a yearly trend where born in 1937=0, accounts for the trend in the difference in smoking behavior between males and females. See note 20 for details on the family income variable. The instrument for education is the induction risk in Vietnam faced by each birth cohort of males multiplied by the risk of being killed in action. The risk of induction is the average yearly number of inductions in Vietnam during the years when that birth cohort was aged 19-22, divided by the size of that birth cohort (See note 23 for the data source). The risk of being killed in action is the ratio between the number of soldiers killed in action in Vietnam in a year and the level of troops during that year (See notes 26 and 27 for details and data sources). Sample sizes vary across specifications as the Vietnam veteran status and the income variable might have been missing.

Figure 1. Prevalence of Smoking in the United States, Age 25 and Above, 1940-2000



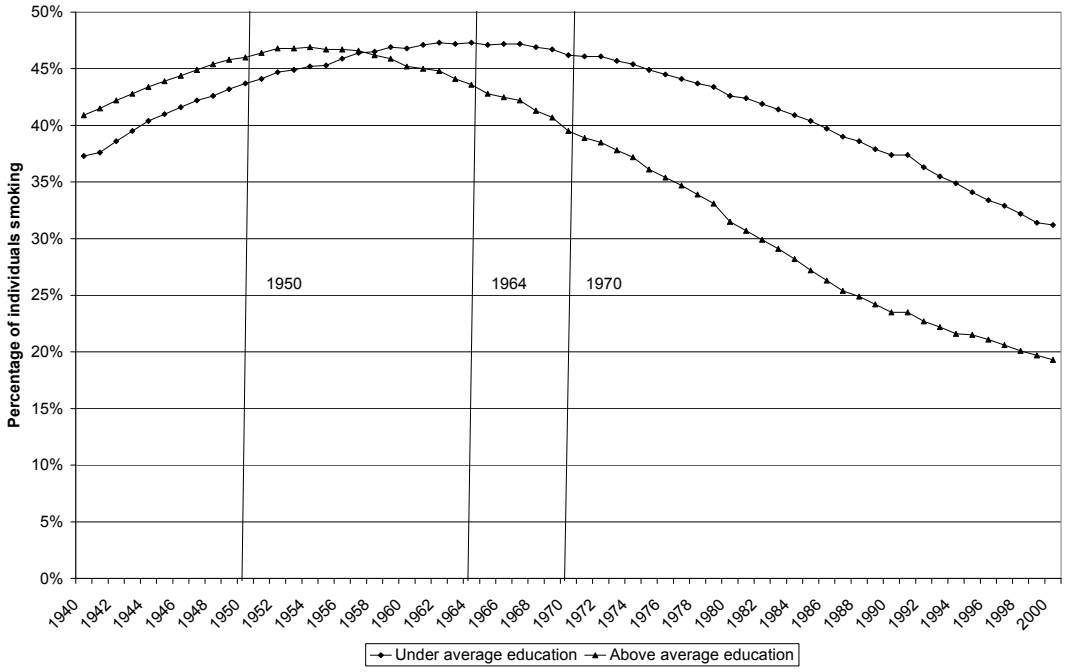
Note: From smoking histories constructed from the 1978, 1979, 1980, 1983, 1985, 1987, 1988, 1990, 1991, 1992, 1994, 1995, 1997, 1998, 1999 and 2000 National Health Interview Surveys. The information about the dangers of smoking diffused gradually: 1950, consensus in medical journals, 1964, first Surgeon General's Report, 1970, clear health warning on packages.

Figure 2. Prevalence of Smoking by Education Category in the United States, Age 25 and above, 1940-2000



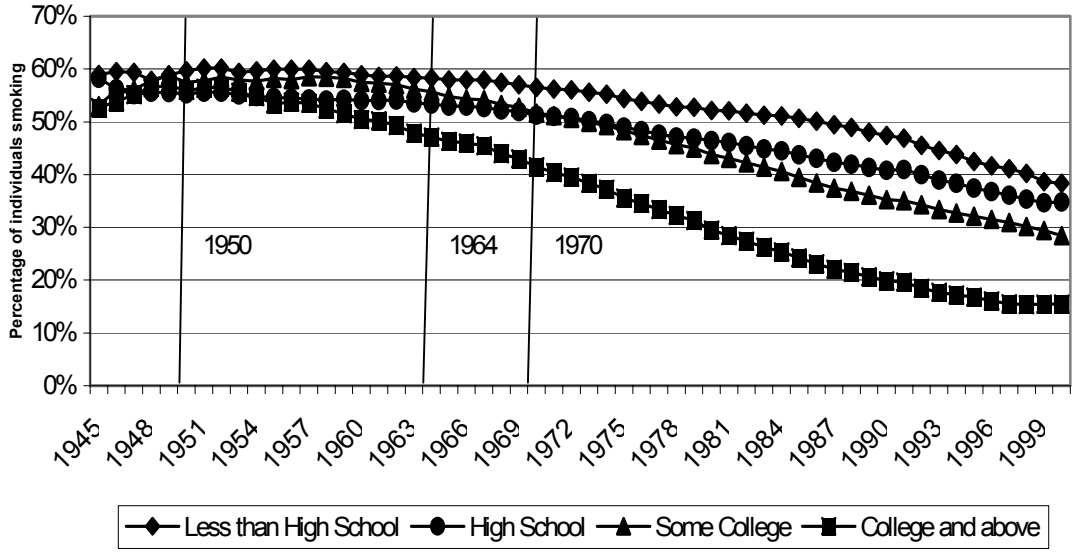
Note: From smoking histories constructed from the 1978, 1979, 1980, 1983, 1985, 1987, 1988, 1990, 1991, 1992, 1994, 1995, 1997, 1998, 1999 and 2000 National Health Interview Surveys. The information about the dangers of smoking diffused gradually: 1950, consensus in medical journals, 1964, first Surgeon General's Report, 1970, clear health warning on packages.

Figure 3. Prevalence of Smoking by Relative Educational Level in the United States, Age 25 and Above, 1940-2000



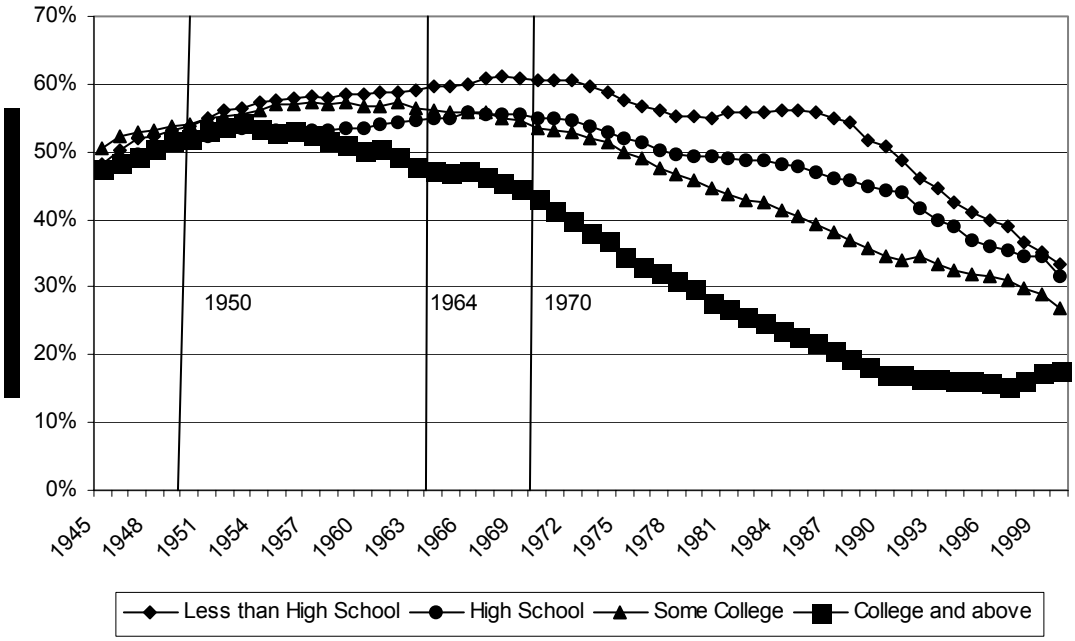
Note: From smoking histories constructed from the 1978, 1979, 1980, 1983, 1985, 1987, 1988, 1990, 1991, 1992, 1994, 1995, 1997, 1998, 1999 and 2000 National Health Interview Surveys. The information about the dangers of smoking diffused gradually: 1950, consensus in medical journals, 1964, first Surgeon General's Report, 1970, clear health warning on packages. Each individual is classified as above or under the average educational achievement in each year for individuals who were aged 25 in that particular year.

Figure 4. Prevalence of Smoking by Education Category in the United States, Age 25 to 60 at the Time of the Survey, 1945-2000



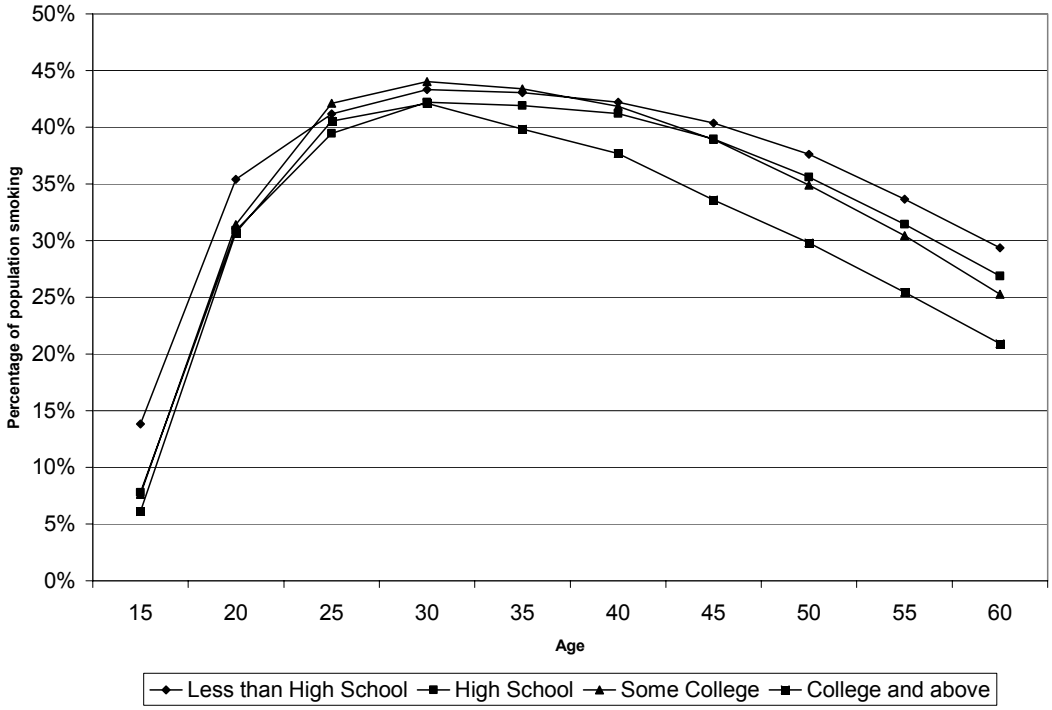
Note: From smoking histories constructed from the 1978, 1979, 1980, 1983, 1985, 1987, 1988, 1990, 1991, 1992, 1994, 1995, 1997, 1998, 1999 and 2000 National Health Interview Surveys. Only individuals aged less than 60 at the time of the interview were selected. The information about the dangers of smoking diffused gradually: 1950, consensus in medical journals, 1964, first Surgeon General's Report, 1970, clear health warning on packages.

Figure 5. Prevalence of Smoking by Education Category in the United States, Age 25-30, 1945-2000



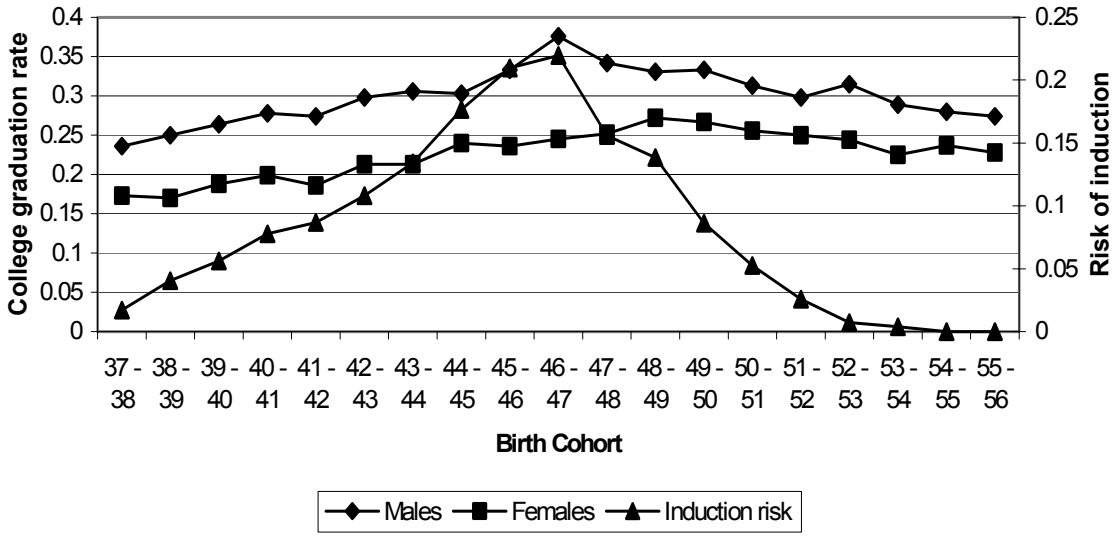
Note: From smoking histories constructed from the 1978, 1979, 1980, 1983, 1985, 1987, 1988, 1990, 1991, 1992, 1994, 1995, 1997, 1998, 1999 and 2000 National Health Interview Surveys. The information about the dangers of smoking diffused gradually: 1950, consensus in medical journals, 1964, first Surgeon General's Report, 1970, clear health warning on packages.

Figure 6. Prevalence of Smoking by Age and Education Category: Cohort Born in 1915-20



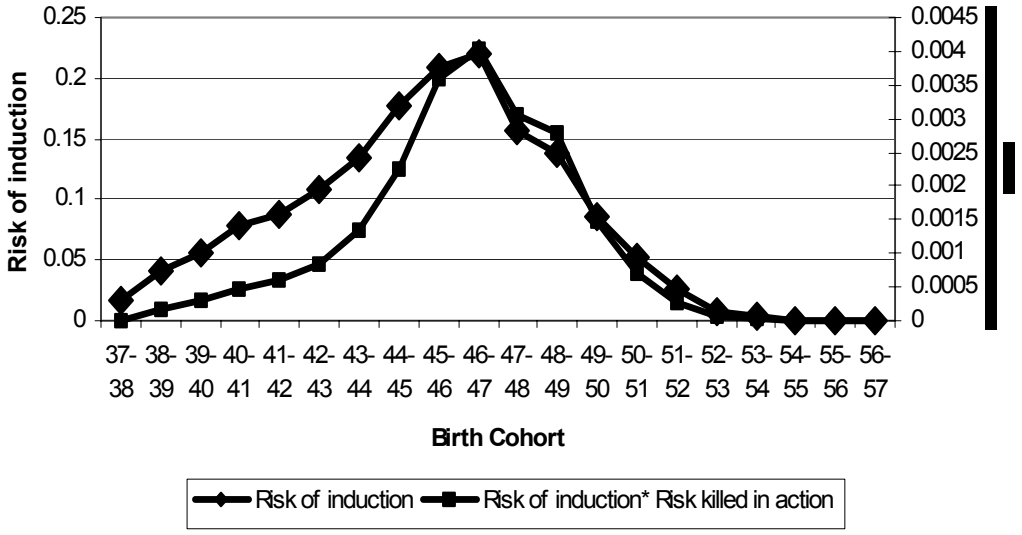
Note: From smoking histories constructed from the 1978, 1979, 1980, 1983, 1985, 1987, 1988, 1990, 1991, 1992, 1994, 1995, 1997, 1998, 1999 and 2000 National Health Interview Surveys.

Figure 7. College Graduation Rates by Gender and Birth Cohort and Risk of Induction in Vietnam, by Birth Cohort



Note: College graduation rates are calculated for individuals born between 1937 and 1956 from the 1983, 1985, 1987, 1988, 1990, 1991, 1992, 1994 and 1995 National Health Interview Surveys. The risk of induction in Vietnam is the average yearly number of inductions in Vietnam during the years when that birth cohort was aged 19-22, divided by the size of that birth cohort, calculated from data in Card and Lemieux [2000] and United States Department of Education [2000].

Figure 8. Risk of Induction in Vietnam and Risk of Induction in Vietnam Multiplied by the Risk of being Killed in Action, by Birth Cohort



Note: The risk of induction in Vietnam is the average yearly number of inductions in Vietnam during the years when that birth cohort was aged 19-22, divided by the size of that birth cohort, calculated from data in Card and Lemieux [2000] and United States Department of Education [2000]. The risk of being killed in action is the ratio between the number of soldiers killed in action in Vietnam in a year and the level of troops during that year (see notes 26 and 27 for details and data sources).